

New approaches to Galaxy Photometric Redshift estimation using Gaussian Process Regression

Michael Way (NASA/Goddard Institute for Space Studies)

Paul Gazis, Jeffrey Scargle (NASA/Ames, Space Sciences Division)

Ashok Srivastava (NASA/Ames Intelligent Systems Division)

Les Foster + Students (San Jose State University)

<http://astrophysics.arc.nasa.gov/~mway/JHU-200904.pdf>

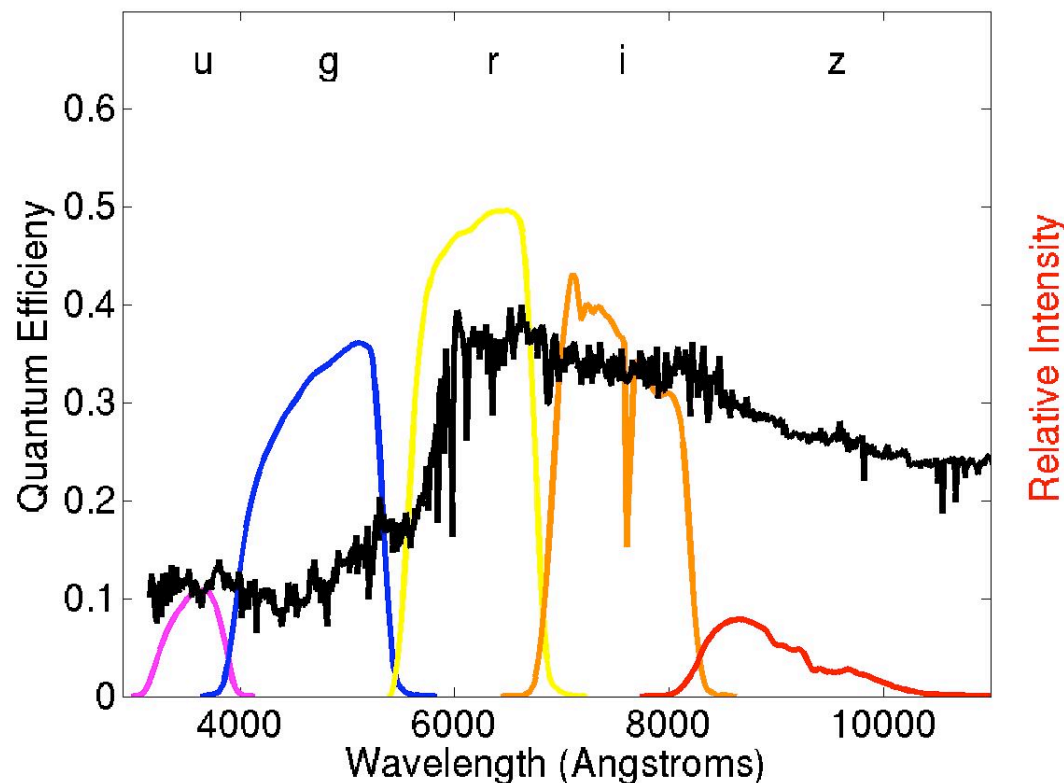
- What are Photometric Redshifts?
- Common training set methods
- What is Gaussian Process Regression?
- Do different kinds of Kernels matter?
- Matrix Inversion Options
- How many galaxies do I need to get a good fit?
- Do SDSS morphological indicators help?
- Do SDSS + 2MASS colors really help?

What are Photometric Redshifts?

Photometric Redshifts: A **rough** estimate of the redshift of a galaxy without having to measure a spectrum.

$$Z_{\text{spec}} = (\lambda_{\text{measured}} - \lambda_{\text{rest}}) / \lambda_{\text{rest}}$$

$$Z_{\text{photo}} = z(C, m)$$

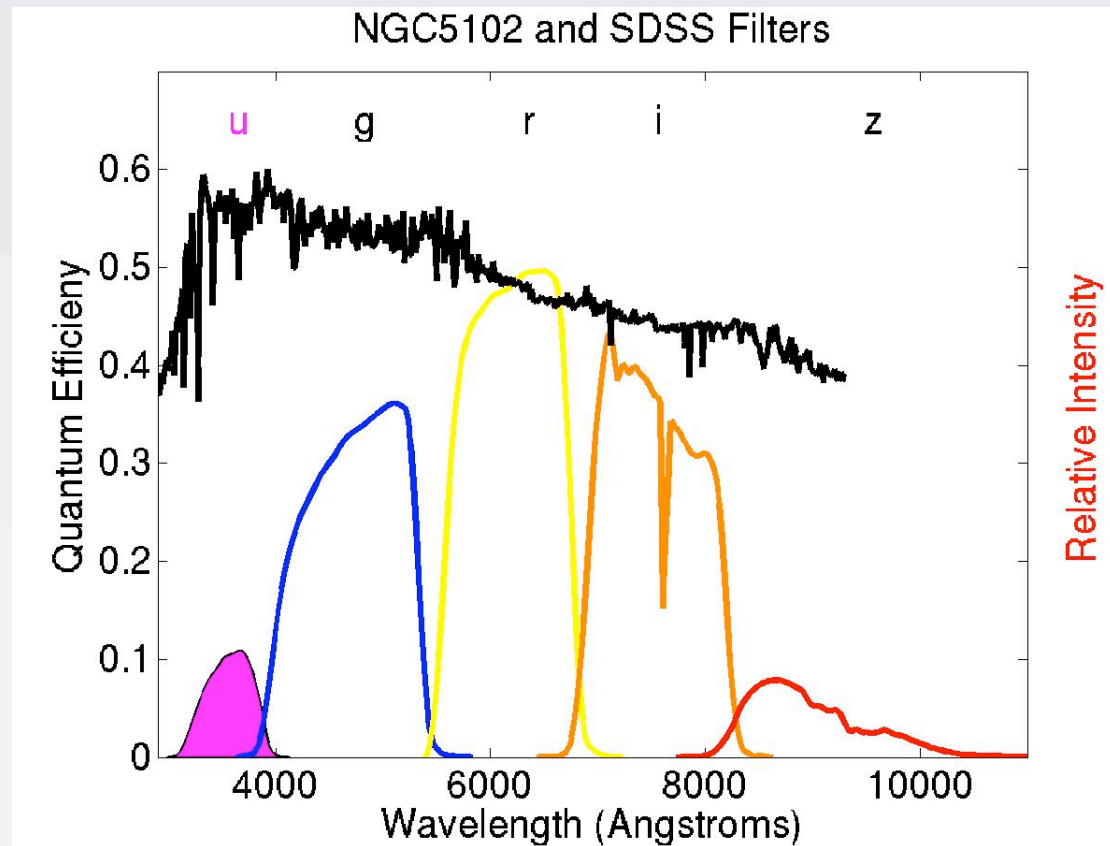


What are Photometric Redshifts?

$$Z_{\text{spec}} = (\lambda_{\text{measured}} - \lambda_{\text{rest}}) / \lambda_{\text{rest}}$$

$$z = 0.0$$

$$z_{\text{photo}} = z(C, m)$$



JHU 2009/04

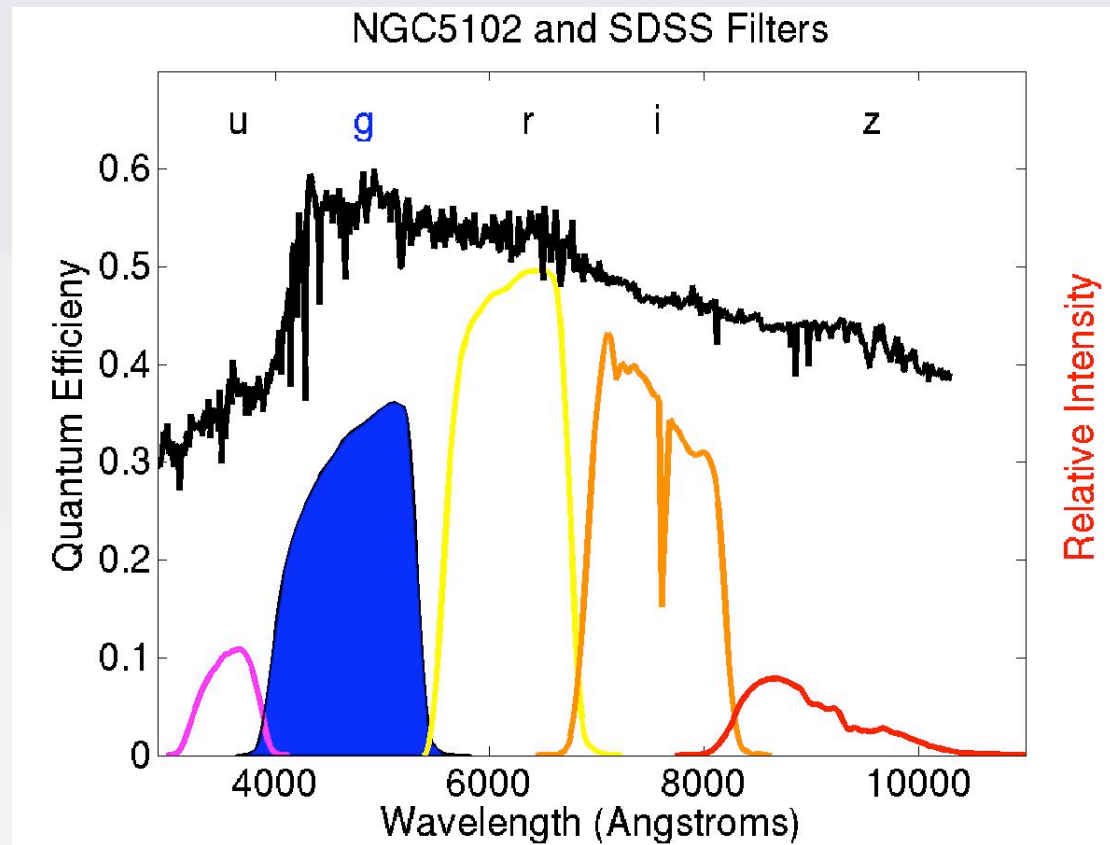


What are Photometric Redshifts?

$$Z_{\text{spec}} = (\lambda_{\text{measured}} - \lambda_{\text{rest}}) / \lambda_{\text{rest}}$$

$$z_{\text{photo}} = z(C, m)$$

$z \sim 0.06$ (18000 km/s)



JHU 2009/04

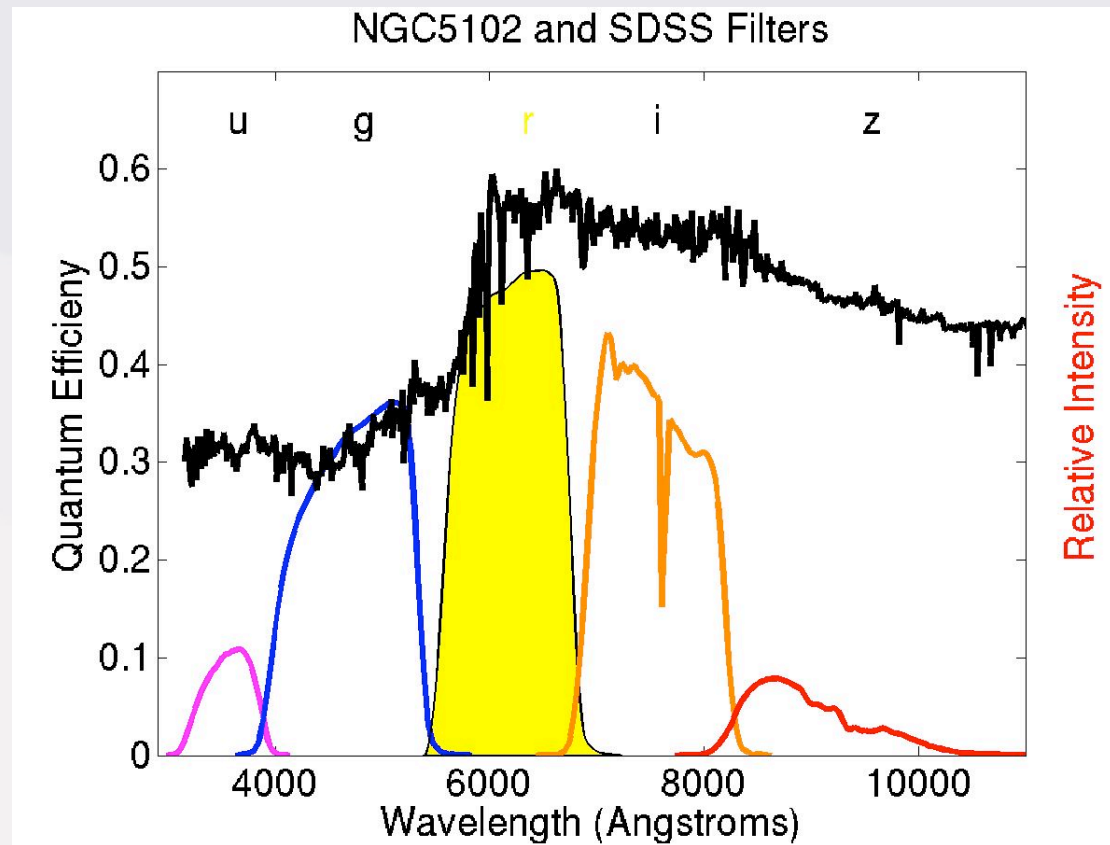


What are Photometric Redshifts?

$$Z_{\text{spec}} = (\lambda_{\text{measured}} - \lambda_{\text{rest}}) / \lambda_{\text{rest}}$$

$$z_{\text{photo}} = z(C, m)$$

$z \sim 0.6$



JHU 2009/04

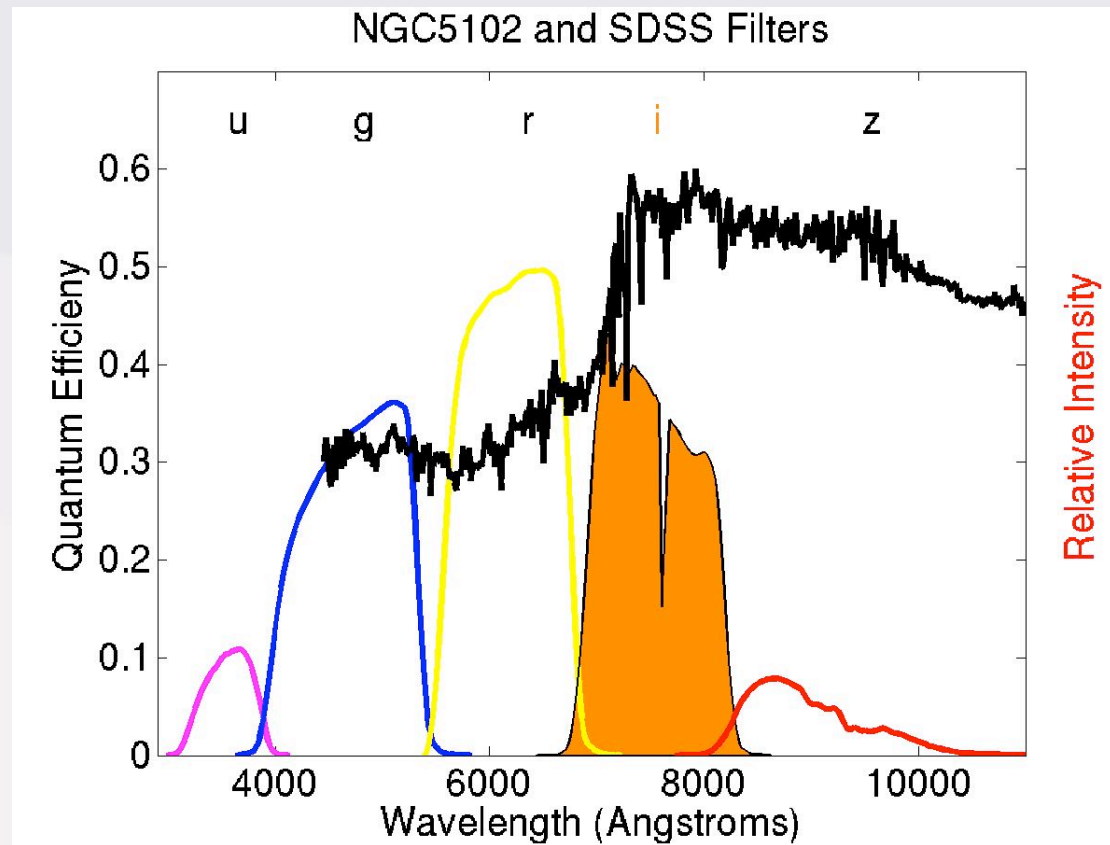


What are Photometric Redshifts?

$$Z_{\text{spec}} = (\lambda_{\text{measured}} - \lambda_{\text{rest}}) / \lambda_{\text{rest}}$$

$$z_{\text{photo}} = z(C, m)$$

$z \sim 0.90$



JHU 2009/04

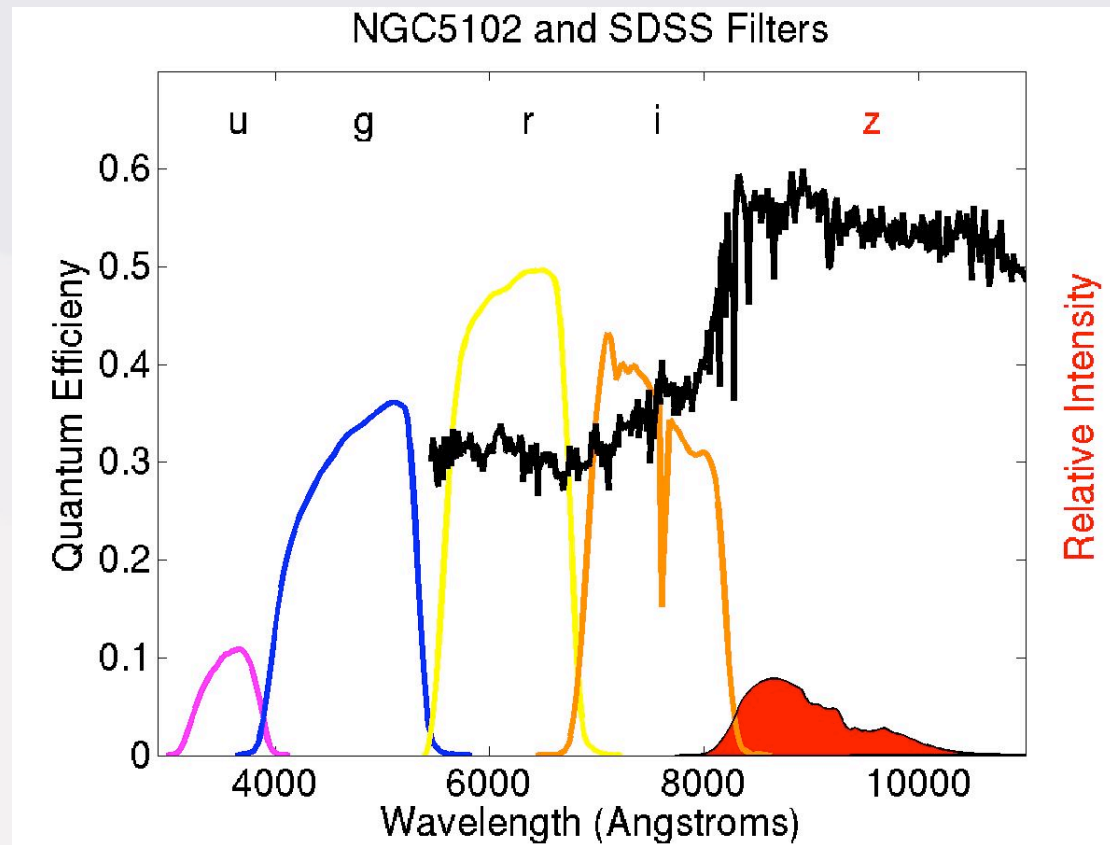


What are Photometric Redshifts?

$$Z_{\text{spec}} = (\lambda_{\text{measured}} - \lambda_{\text{rest}}) / \lambda_{\text{rest}}$$

$$z_{\text{photo}} = z(C, m)$$

$z \sim 1.10$



JHU 2009/04





Photo-z methods

1.) Spectral Energy Distribution (SED) Fitting:

- model based approach
- uses redshifts derived from spectra of artificial galaxies (e.g. Bruzual & Charlot)

2.) Training-Set methods:

- **empirical approach**
- **uses *spectroscopic* redshifts from a sub-sample of galaxies with the same band-pass filters**

Photo-z The Empirical Approach

Training Set Methods need a sub-sample of Galaxies:

- of known spectroscopic redshift
- with a comparable range of **magnitudes** (u g r i z) to our Photometric survey objects
For the SDSS MGS that is $r < 17.77$ (NOT $17.77 < r < 22$)
- These will be our “Training Samples”





“Training Set” Methods

Galaxy Photometric Redshift Prediction History

- Linear Regression was first tried in the 1960s
- Quadratic & Cubic Regression (1970s)
- Polynomial Regression (1980s)
- Neural Networks (1990s)
- Kd Trees & Bayesian Classification Approaches (1990s)
- Support Vector Machines & GP Regression (2000s)

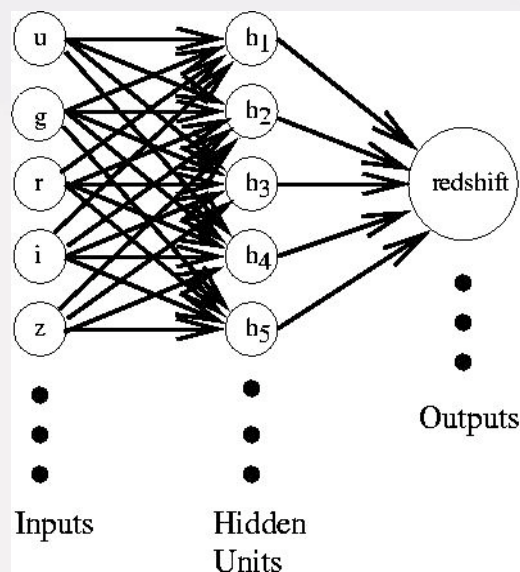


Gaussian Process Regression fitting

Gaussian Process Regression \Leftrightarrow Kernel Methods

Kernel Methods have replaced Neural Networks in the Machine Learning literature

WHY?: given a large # of hidden units \Rightarrow GP (Neal 1996).



$$h_n > 100$$

$\rightarrow \rightarrow \rightarrow \rightarrow \rightarrow$

JHU 2009/04



Kernel Methods - Gaussian Process Regression

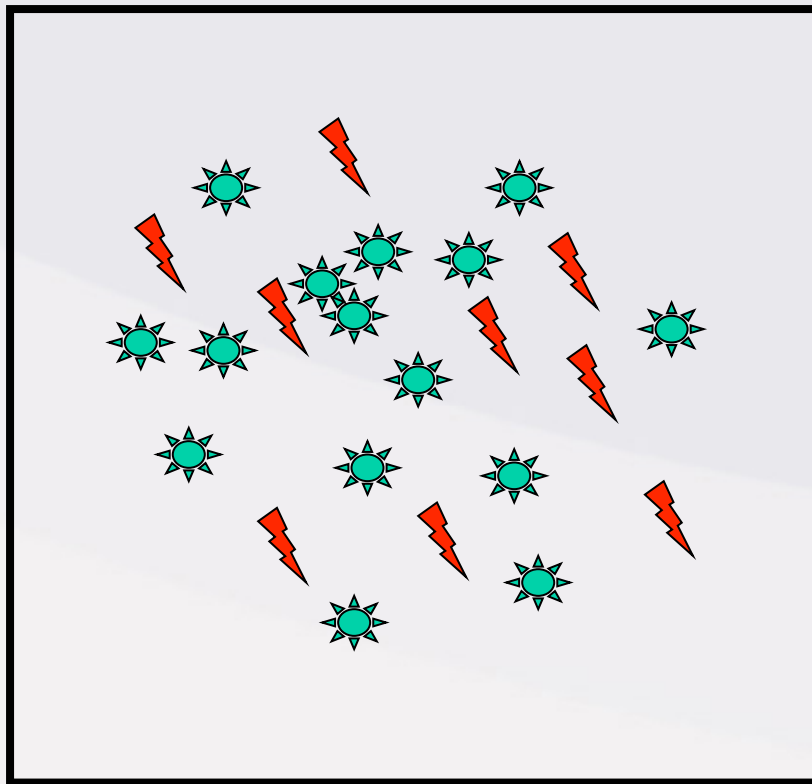
GP regression builds a linear model in a very high dimensional *parameter space* (“feature space” \rightarrow Hilbert space).

- One can map the data using a function $F(x)$ [kernel] into this high (or infinite) dimensional *parameter space* where one can perform linear operations.



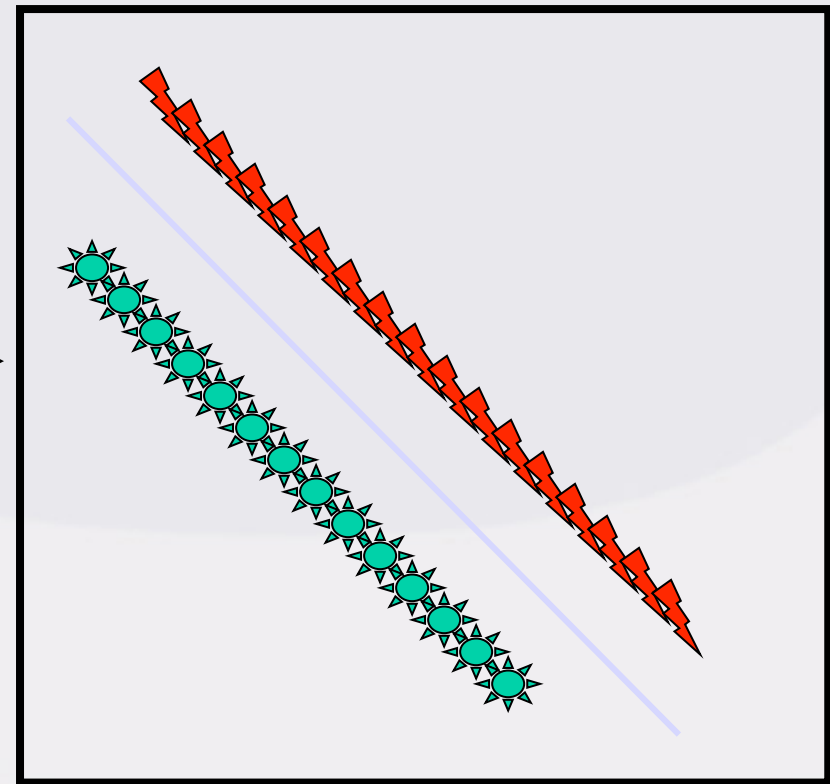
The value of kernels

Original Data without Kernel



Data in original space: highly complex decision boundaries.

Mapped Data using Kernel



Data in high dimensional feature space after mapping through $F(x)$ can yield simple decision boundaries.



GP Regression: Advantages

GP Advantages:

- Small input data training samples
(good for higher redshifts?) yet low errors
- Over fitting is eliminated by use of proper priors
- Realistic estimation of individual redshift errors

GP Regression: Problems?

GP Disadvantages:

- Possibly large CPU time requirements
 - The Kernel (Covariance Matrix) **can** be large:
 $K = (\lambda^2 I + XX^T)^2$ if $X = 5 \times 180,000$ (our case) then
 K is a matrix $180,000 \times 180,000$ and we have:

$$y^* = K^* (\lambda^2 I + K)^{-1} y$$

- Need to invert this large K matrix
 - $O(N^3)$ operation, $O(N^2)$ memory
- Kernel Selection is ambiguous?





GP: Which Kernel??

Kernel Selection: Pick a transfer/covariance function

Matern Class Fcn

$$k(r) = \frac{2^{l-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu r}}{l} \right)^\nu J_\nu \left(\frac{\sqrt{2\nu r}}{l} \right) \quad \nu \rightarrow \infty$$

Radial Basis Fcn

$$k(r) = \exp\left(-\frac{r^2}{2l^2}\right)$$

Rational Quadratic Polynomial

$$k_{RQ}(r) = 1 + \left(\frac{r^2}{2\alpha l^2} \right)^{-\alpha}$$

$$k(x, x') = \left(\sigma_o^2 + x^T \sum_p x' \right)^p$$

Neural Nets

$$k_{NN}(x, x') = \frac{2}{\pi} \sin^{-1} \left(\frac{2x^T \Sigma x'}{\sqrt{(1 + 2x^T \Sigma x)(1 + 2x'^T \Sigma x')}} \right)$$



GP Matrix Inversion

Matrix Inversion: 3 options

Option 1: Take a random sample of ~ 1000 galaxies & invert that while bootstrapping n times from full sample (Paper I)

- Advantages
 - Can run on a 32bit computer
 - Doesn't take too long: $O(N^3)$ operation
 - Doesn't take up too much memory $O(N^2)$
- Disadvantages
 - Accuracy suffers – we don't sample enough galaxies/SEDs



GP Matrix Inversion

Matrix Inversion: 3 options

Option 2: Use a 64 bit SSI computer

- Advantages
 - Accuracy – we invert the full matrix using all sample galaxies
- Disadvantages
 - Takes a VERY long time: $O(N^3)$ operation
 - We need a lot of memory: $O(N^2)$
 - Hard to get access to such a computer for such a long time
 - e.g. Mac Pro: 64 bit, 4 cpu, 16GB of RAM, max is $\sim 20000 \times 20000$ in Matlab



GP Matrix Inversion

Matrix Inversion: 3 options

Option 3: Low-rank matrix approximations: Subset of Regressors, Cholesky Decomposition, Projected Process Approximation, etc.

(Paper II: <https://dashlink.arc.nasa.gov/algorithm/stablegp>)

- Advantages
 - Accuracy – we invert much more of the full matrix
 - Doesn't take too long: $O(N^3)$ operation (dependent upon rank=N)
 - Doesn't take up too much memory $O(N^2)$
- Disadvantages
 - Hard to know how it compares to full matrix inversion

Results: Other authors

Method Name	σ_{rms}	Dataset ¹	Inputs ²	Source
CWW	0.0666	SDSS-EDR	ugriz	Csabai et al. (2003)
Bruzual-Charlot	0.0552	SDSS-EDR	ugriz	Csabai et al. (2003)
ClassX	0.0340	SDSS-DR2	ugriz	Suchkov et al. (2005)
Polynomial	0.0318	SDSS-EDR	ugriz	Csabai et al. (2003)
Support Vector Machine	0.0270	SDSS-DR2	ugriz	Wadadekar (2005)
Kd-tree	0.0254	SDSS-EDR	ugriz	Csabai et al. (2003)
Support Vector Machine	0.0230	SDSS-DR2	ugriz+r50+r90	Wadadekar (2005)
Artificial Neural Network	0.0229	SDSS-DR1	ugriz	Collister & Lahav (2004)



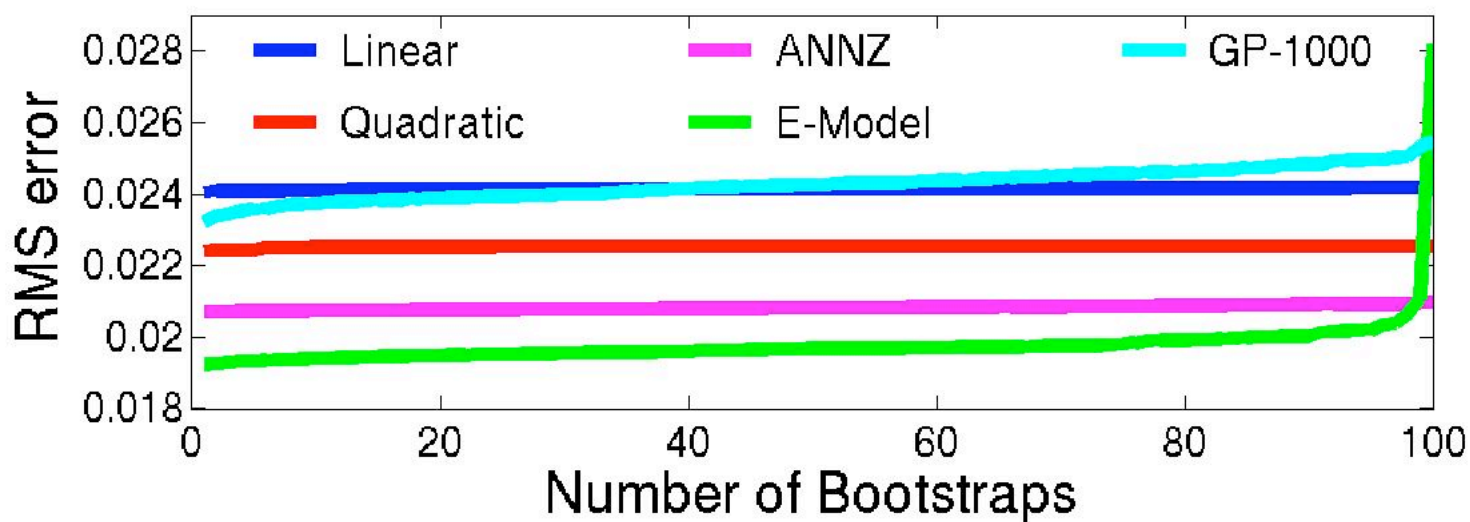
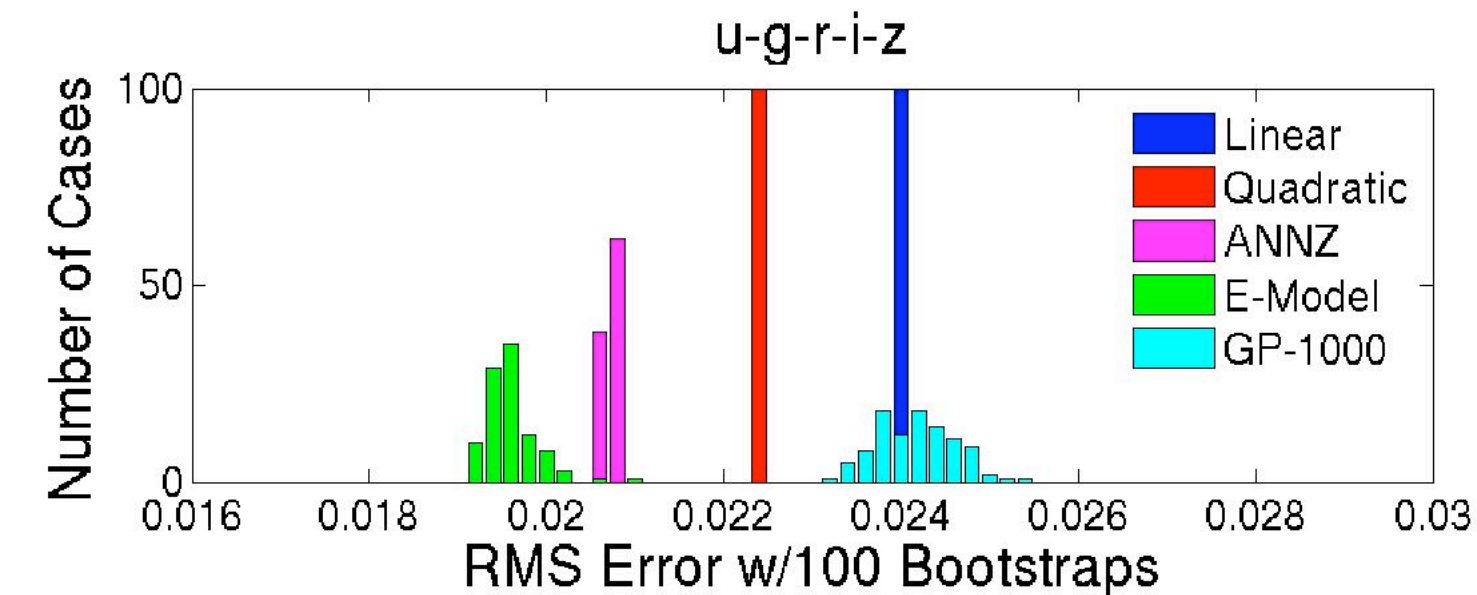


GP Regression (Results)

Results: SDSS (DR3) Main Galaxy Sample

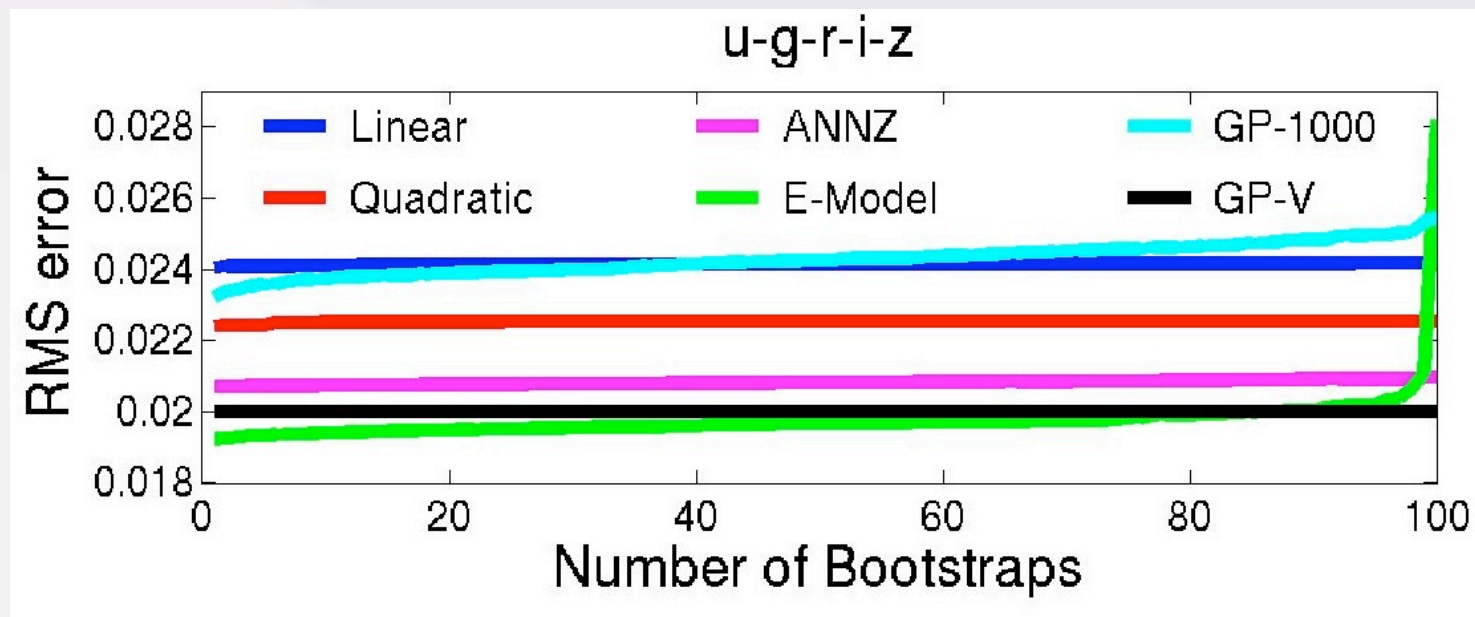
- Paper I: Compared linear, quadratic, Neural Networks and GPs on the SDSS-DR3
 - With ONLY 1000 samples GPs performed well compared to the other methods
- Paper II: Low-rank matrix inversion approximations with more appropriate Kernel
 - GPs performed better than all other methods to date

Paper I Results: Comparing Methods



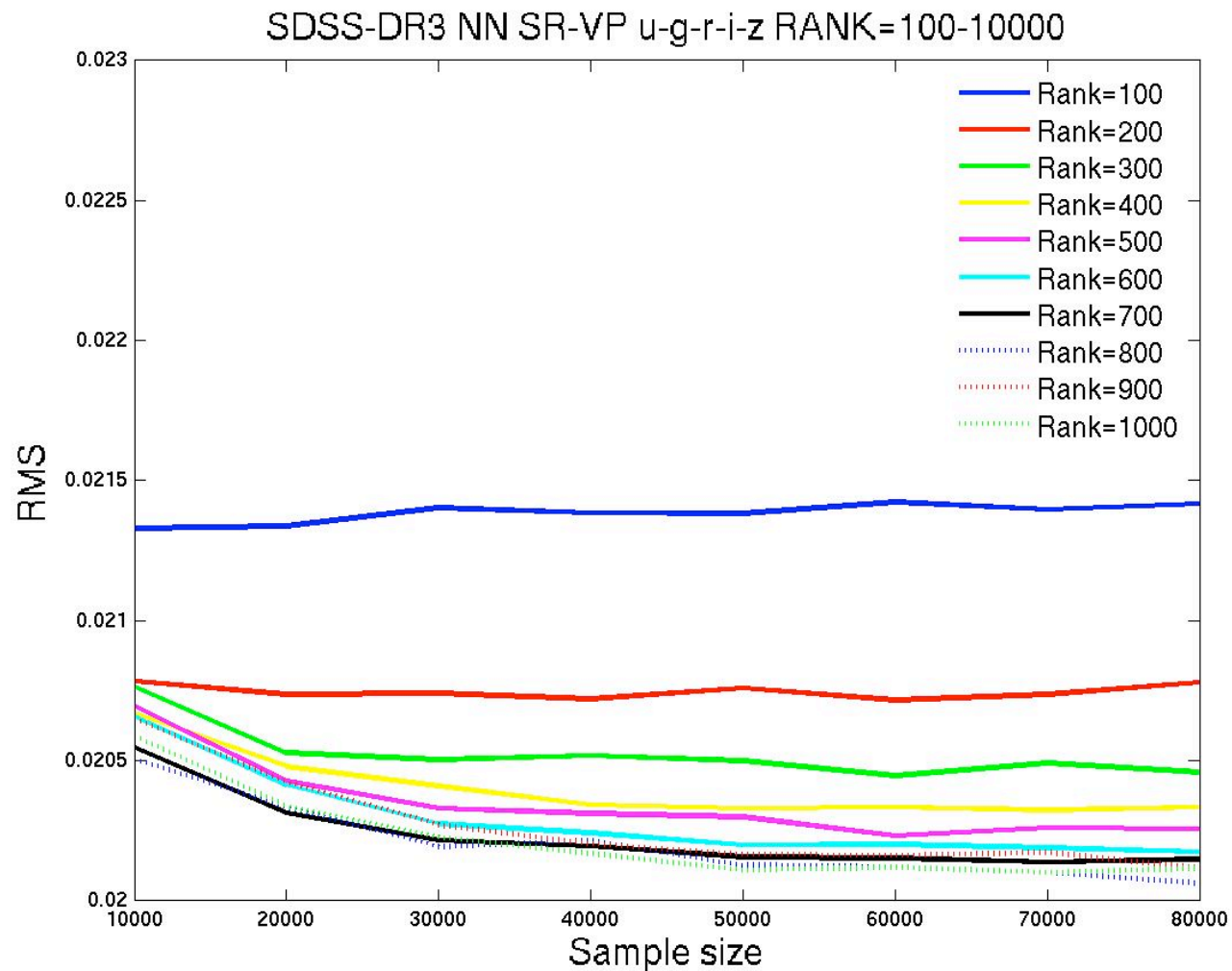
New Results: Paper II

- GPR with rank=1000 : V-method : 36,000 samples : Polynomial Kernel
- Better results possible using VP method & NN Kernels



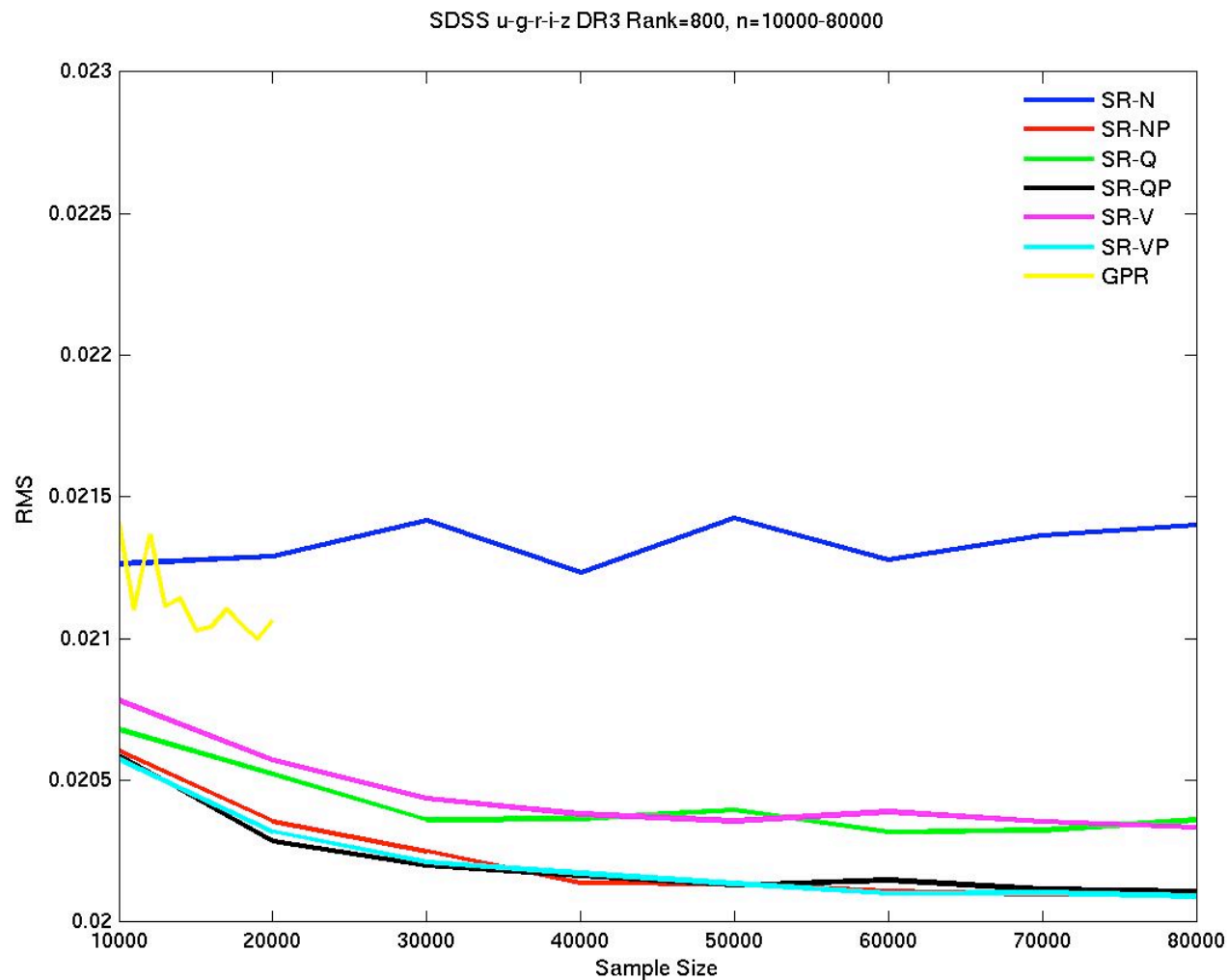
Rank vs Sample Size

Near optimal is ~40,000 samples, Rank=800



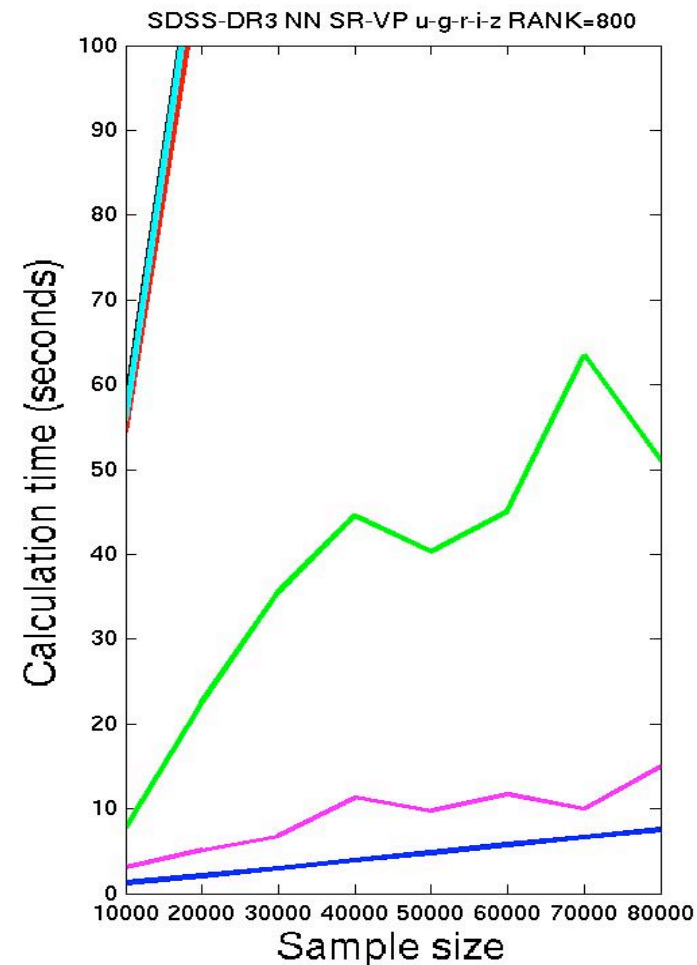
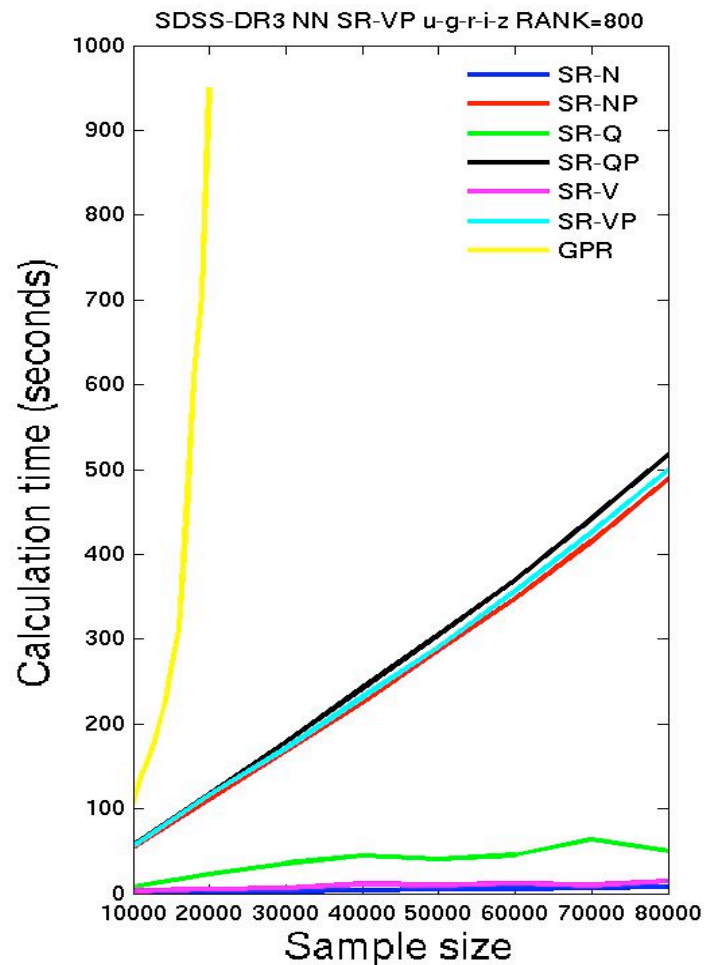
Rank Reduction Methods

Comparing Matrix Inversion Techniques. Why SR-VP ?



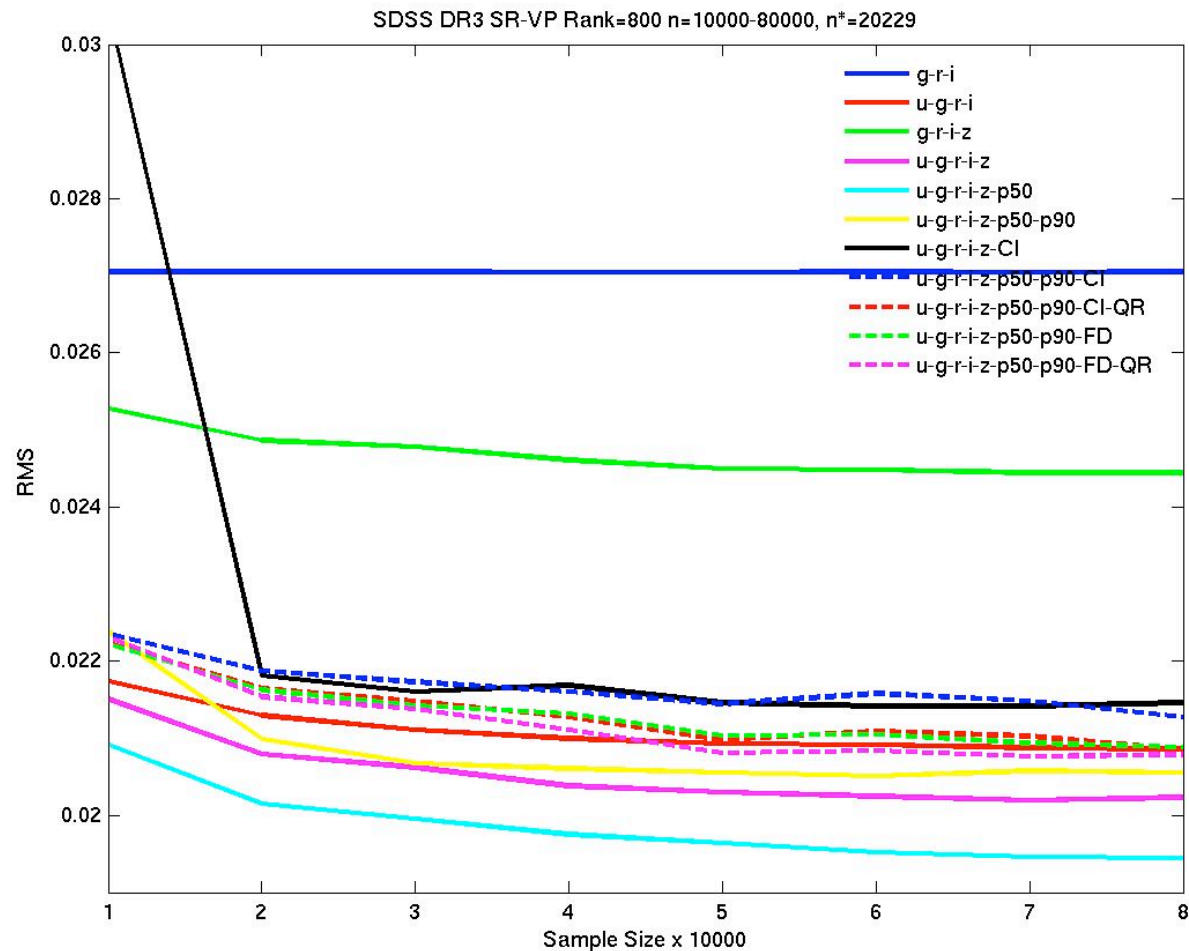
Calculation Time?!

Matrix Inversion: that $O(N^3)$ business?



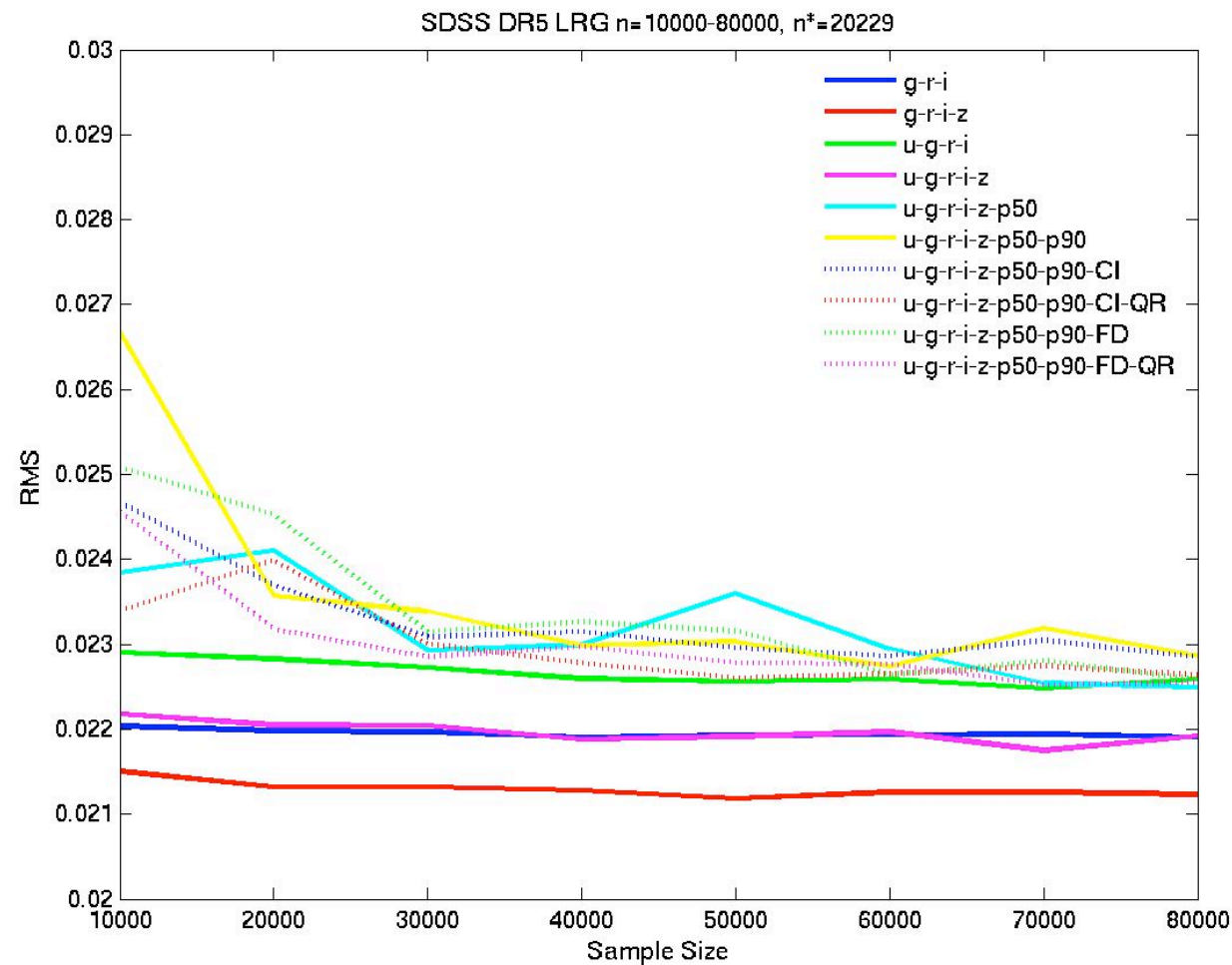
Does SDSS morphology help?

SDSS-DR3 Main Galaxy Sample



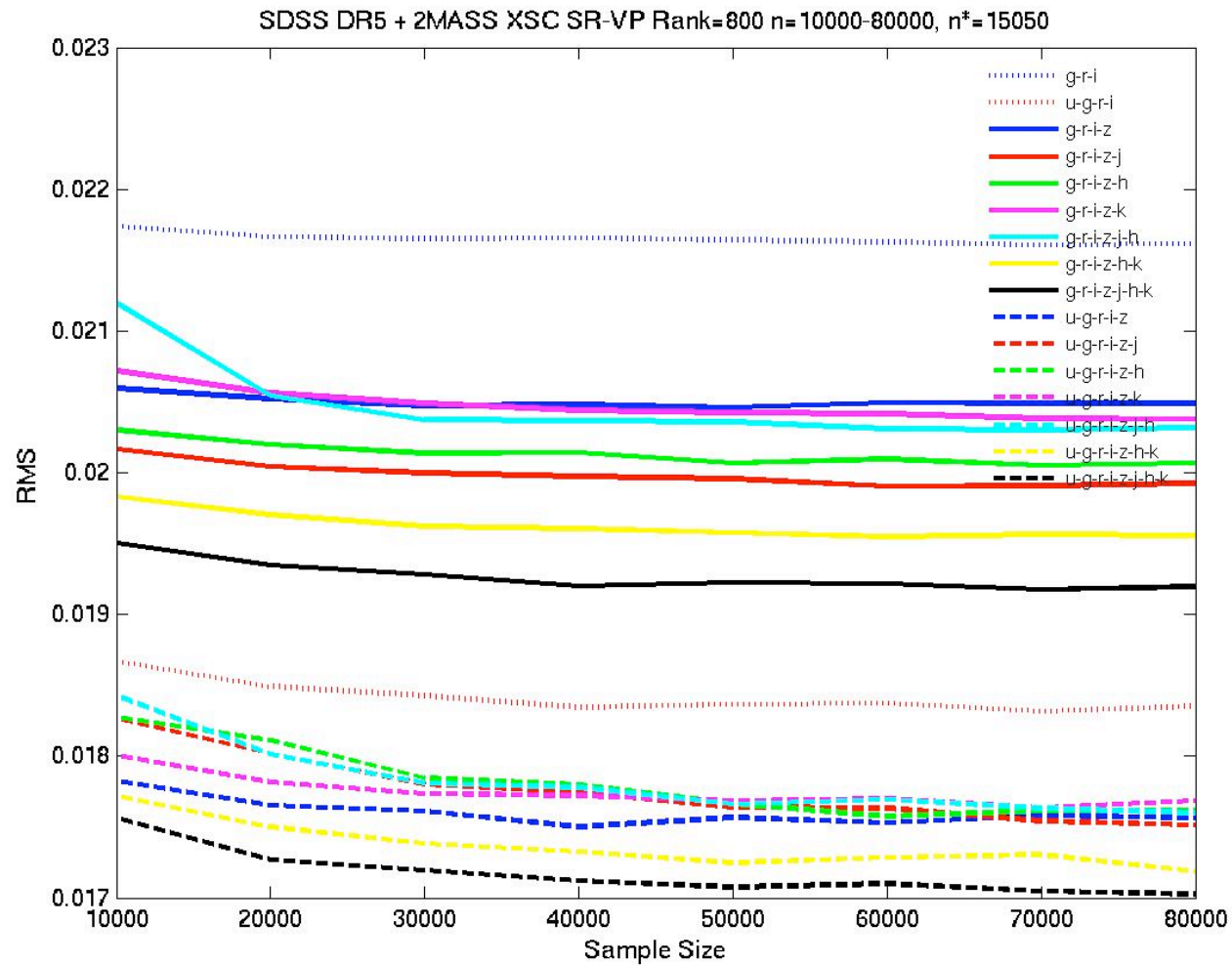
Does SDSS morphology help?

SDSS-DR5 Luminous Red Galaxy Sample



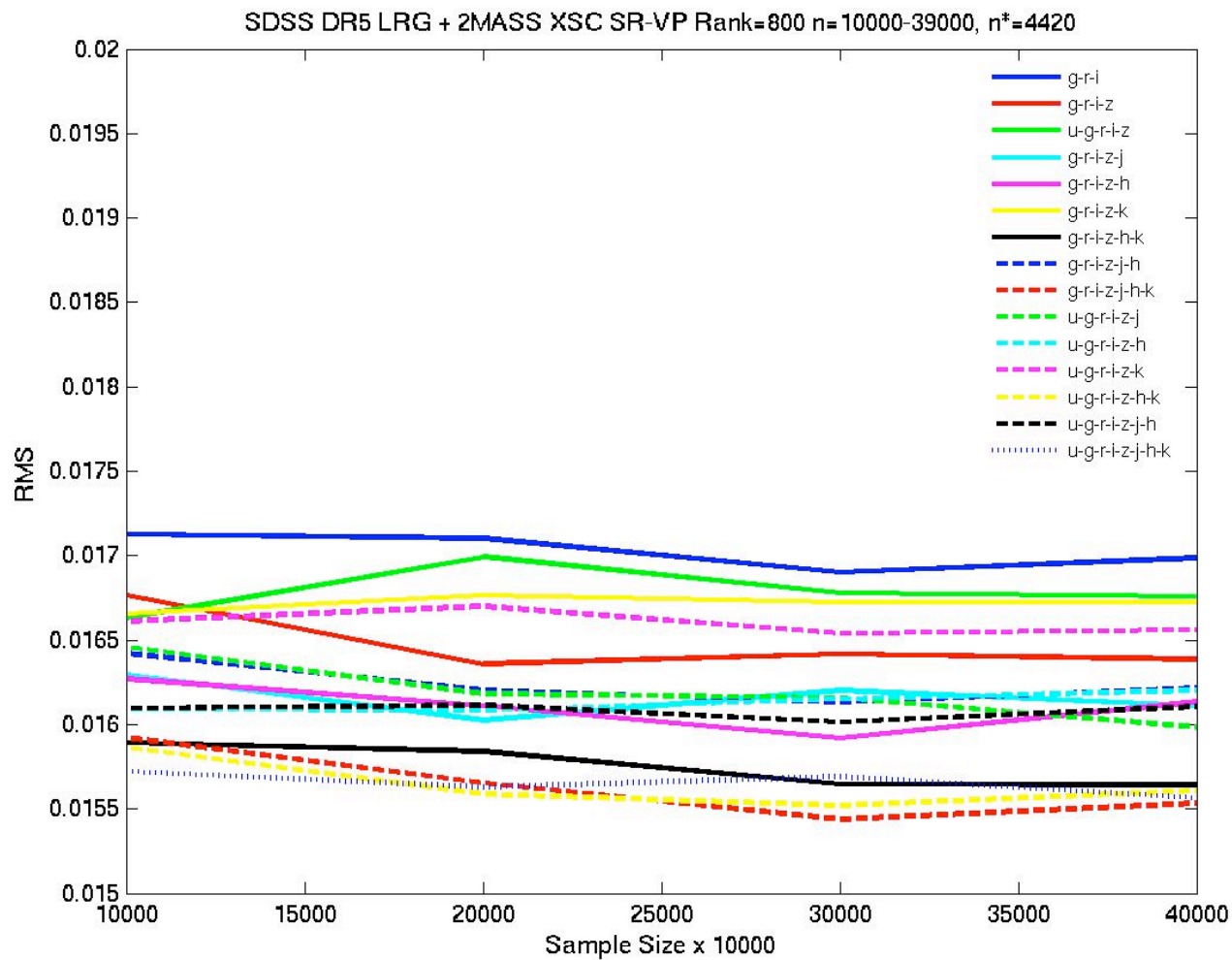
SDSS-MGS + 2MASS xsc

SDSS-DR5 MGS + 2MASS



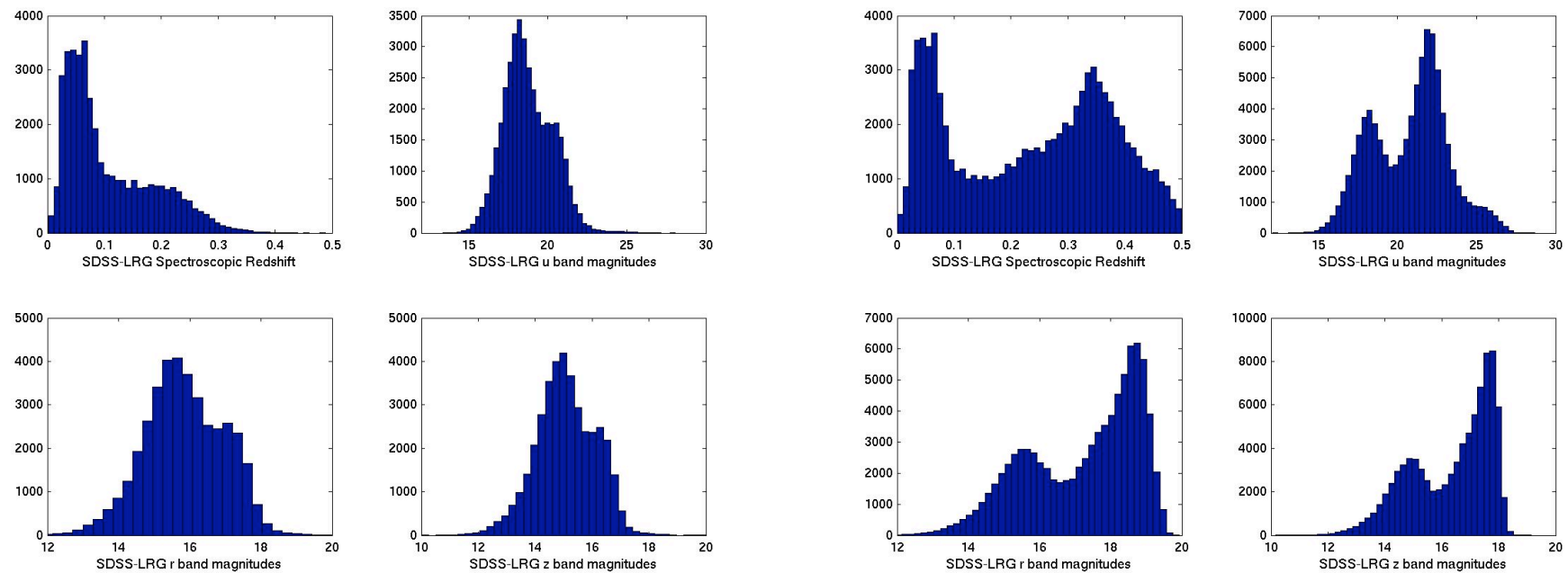
SDSS-LRG + 2MASS xsc

SDSS-DR5 LRG + 2MASS



SDSS + 2MASS xsc

SDSS only magnitudes are suddenly better?



2MASS+SDSS/LRG-DR5

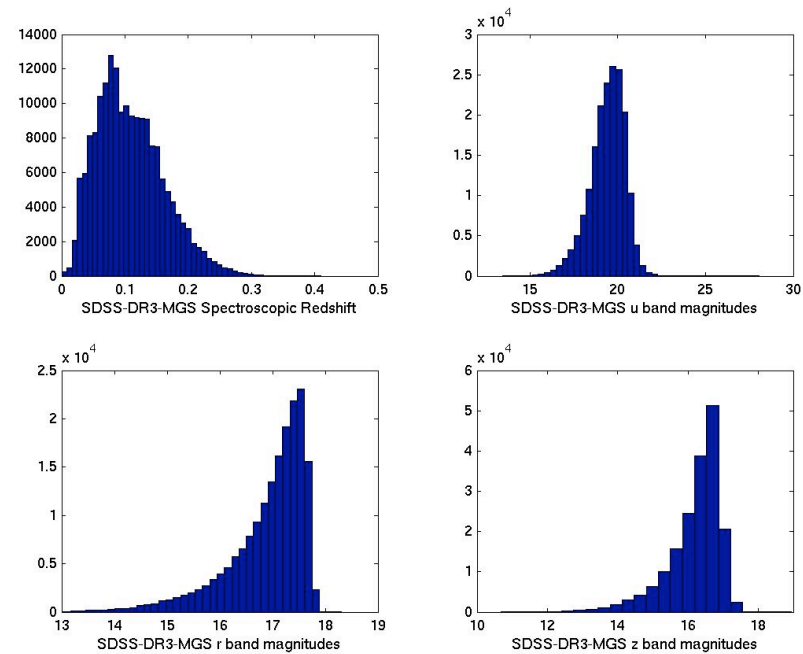
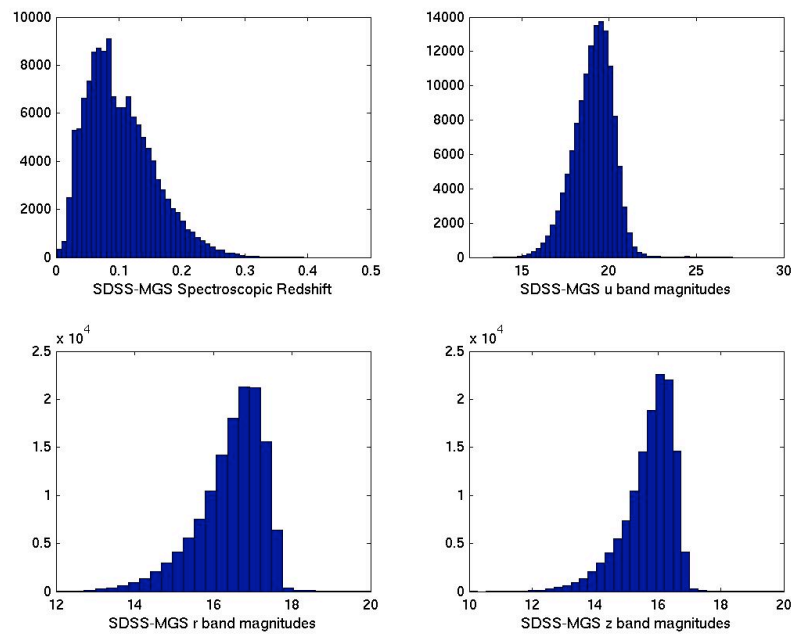
SDSS/LRG-DR5 only

JHU 2009/04



SDSS + 2MASS xsc

SDSS only magnitudes are suddenly better?



2MASS/SDSS-DR5 match

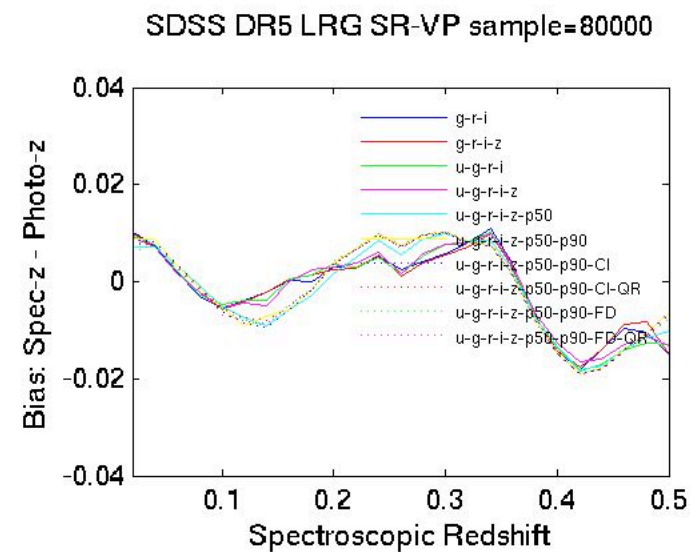
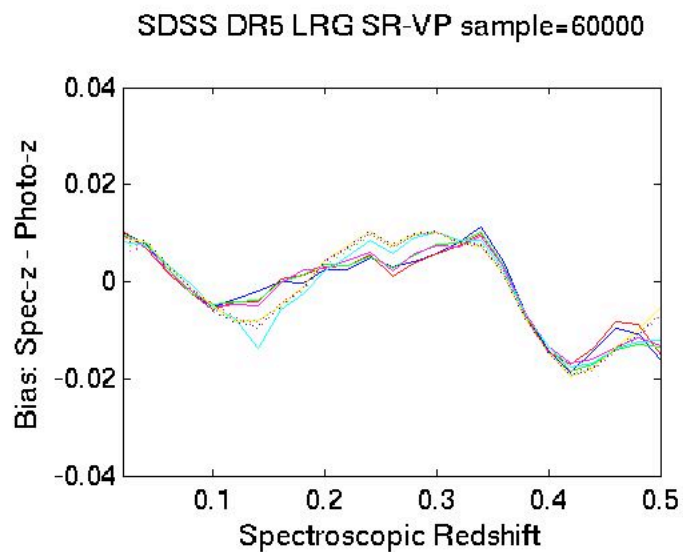
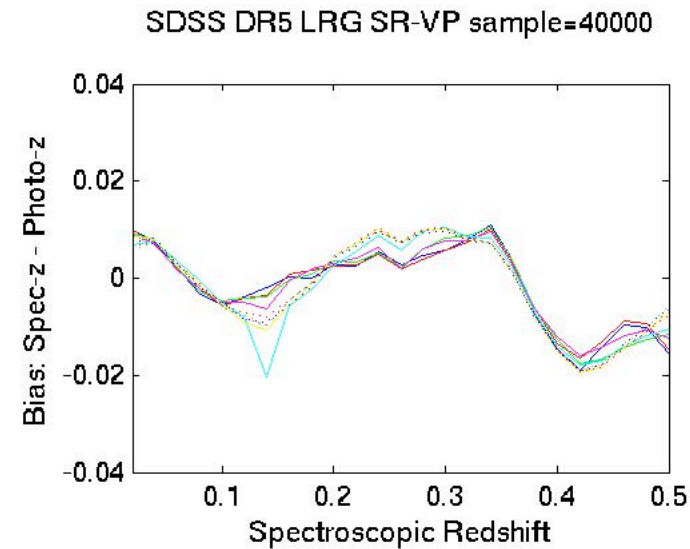
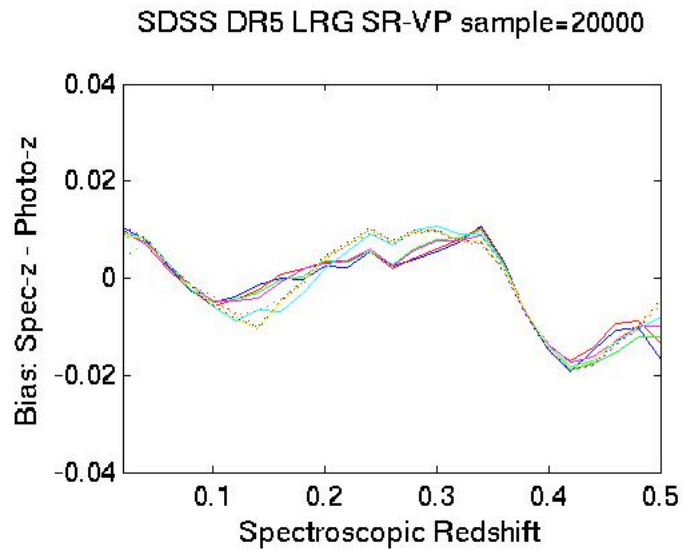
SDSS-DR5 only



Results?

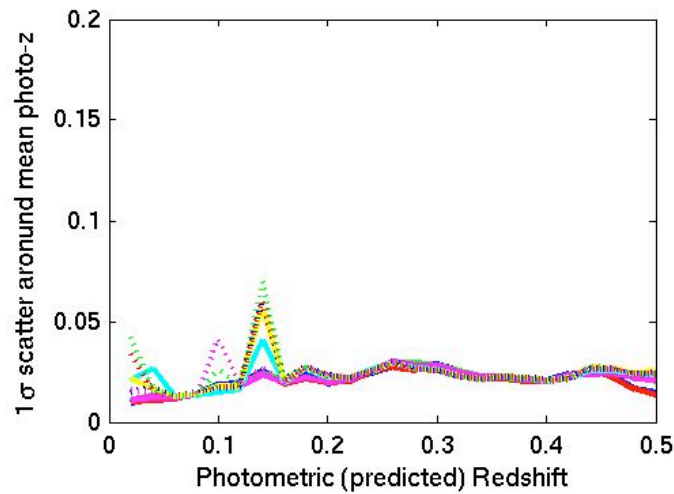
- GPR is now faster & better than most others
- ~40,000 objects are required for optimal results when using the SDSS-MGS, while LRG sample is good at 10,000
- Morphology does not *generally* increase accuracy of photo-z estimation *with GPR*
- Additional Near IR filters (2MASS) increase accuracy, but ...
- Galaxy classification helps: MGS vs LRG

SDSS-DR5/LRG GPR

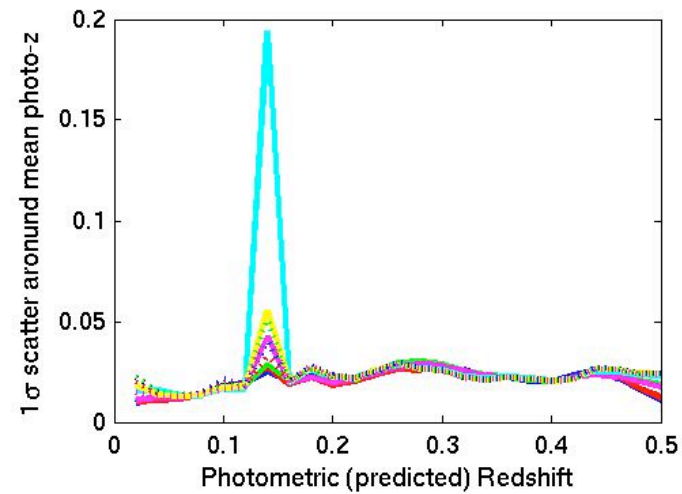


SDSS-DR5/LRG GPR

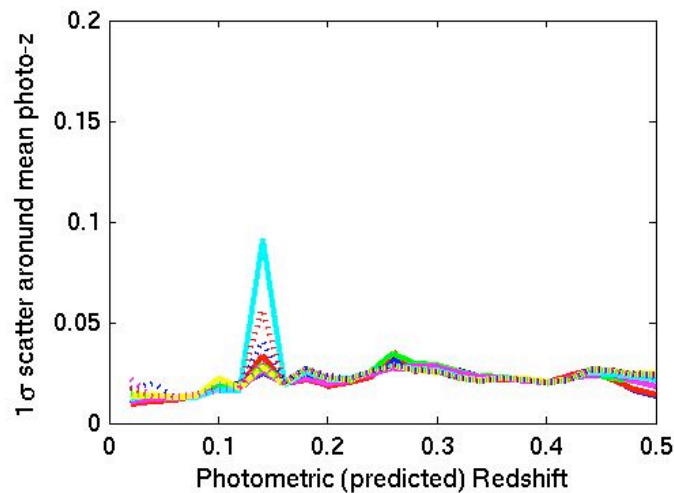
SDSS DR5 LRG SR-VP sample=20000



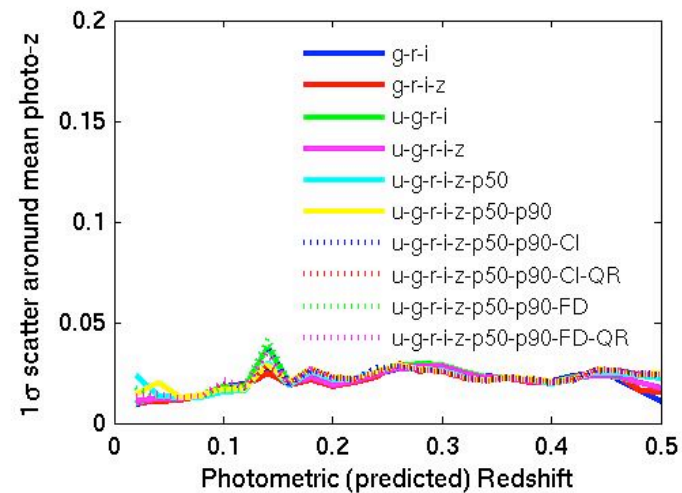
SDSS DR5 LRG SR-VP sample=40000



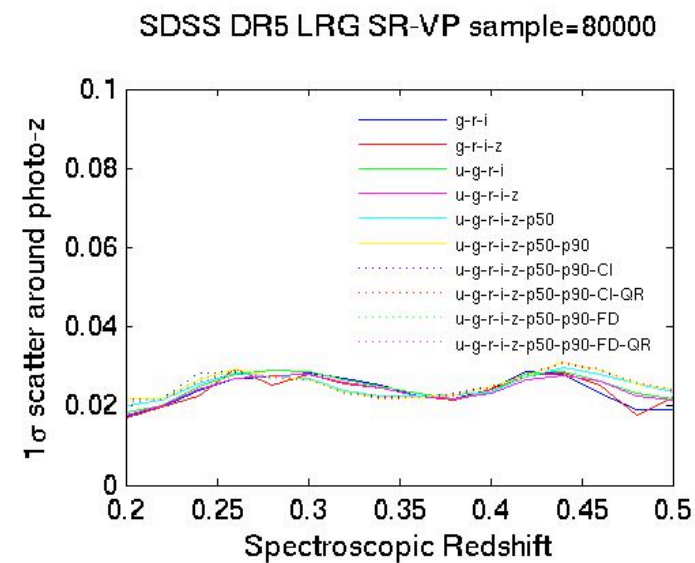
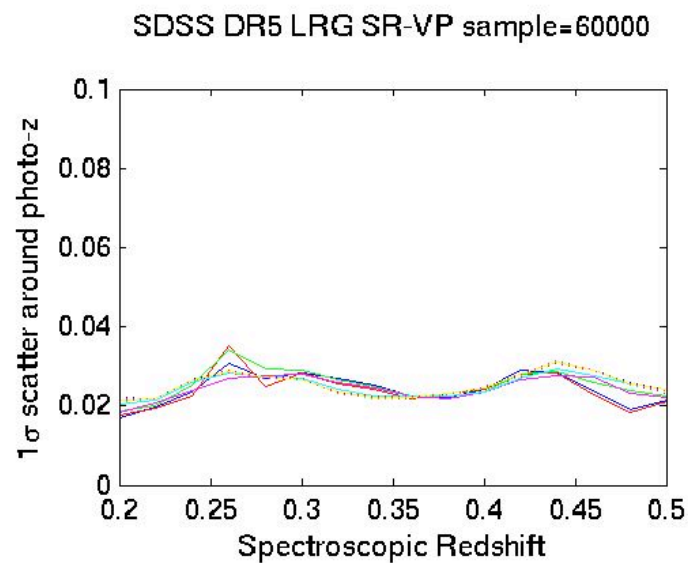
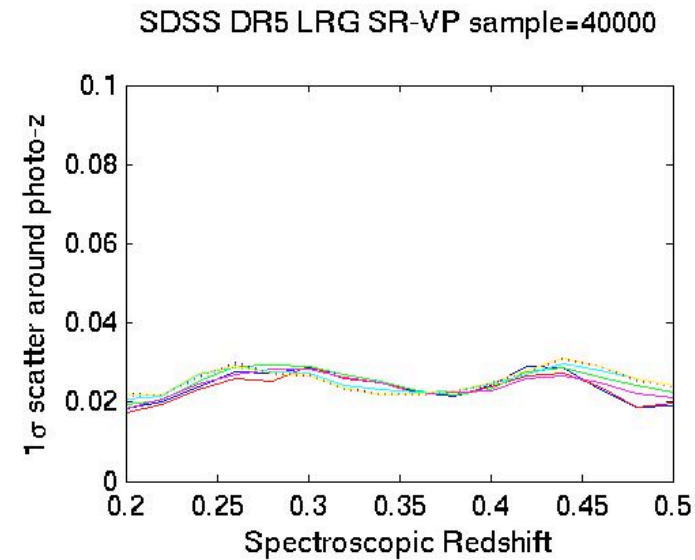
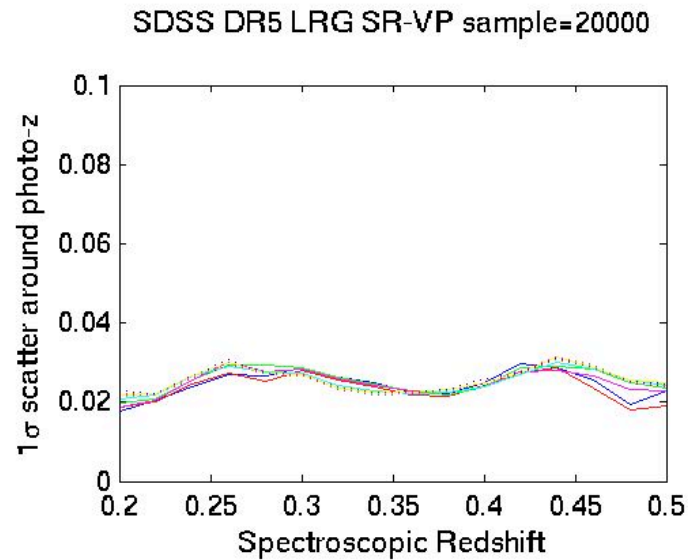
SDSS DR5 LRG SR-VP sample=60000



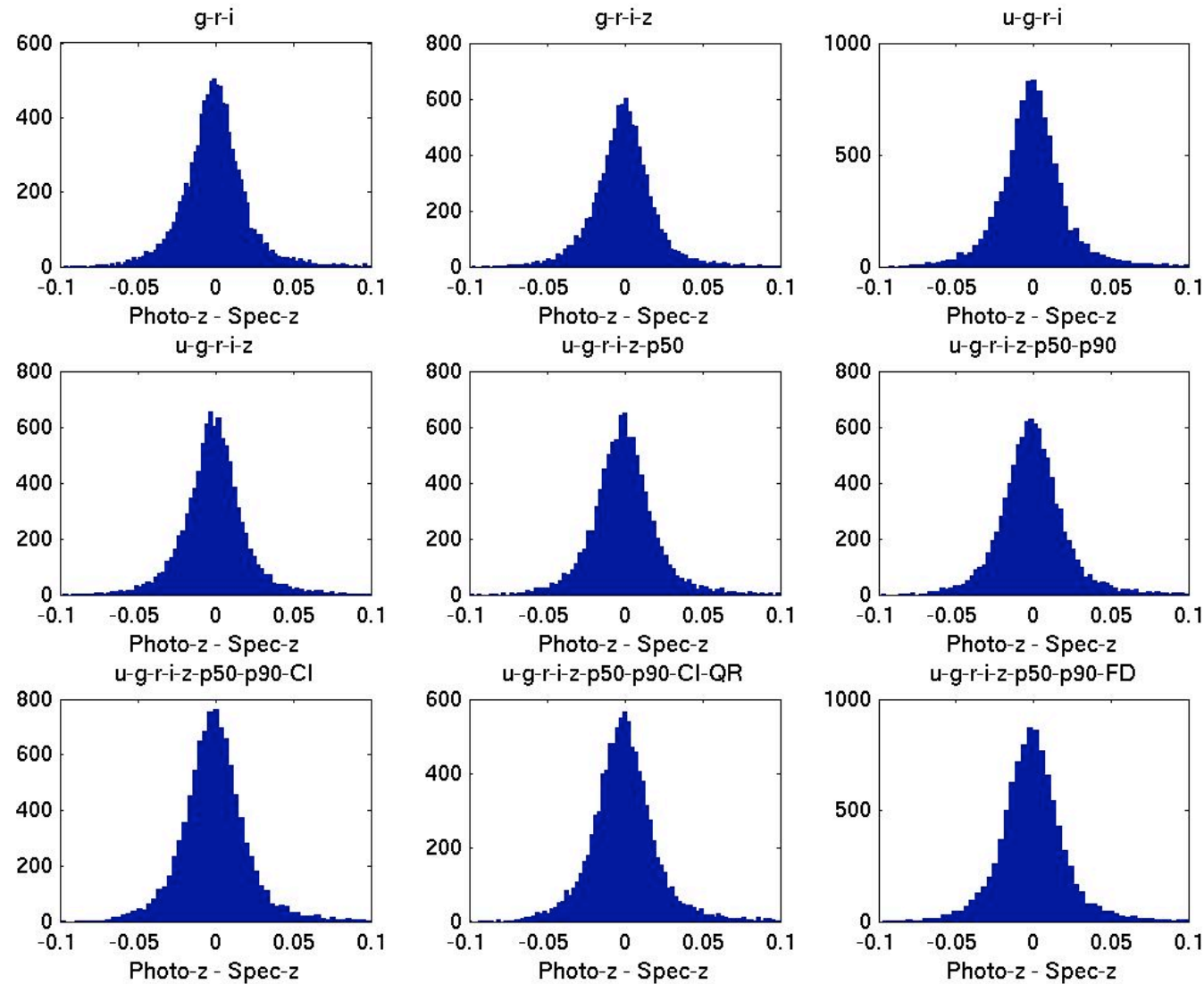
SDSS DR5 LRG SR-VP sample=80000



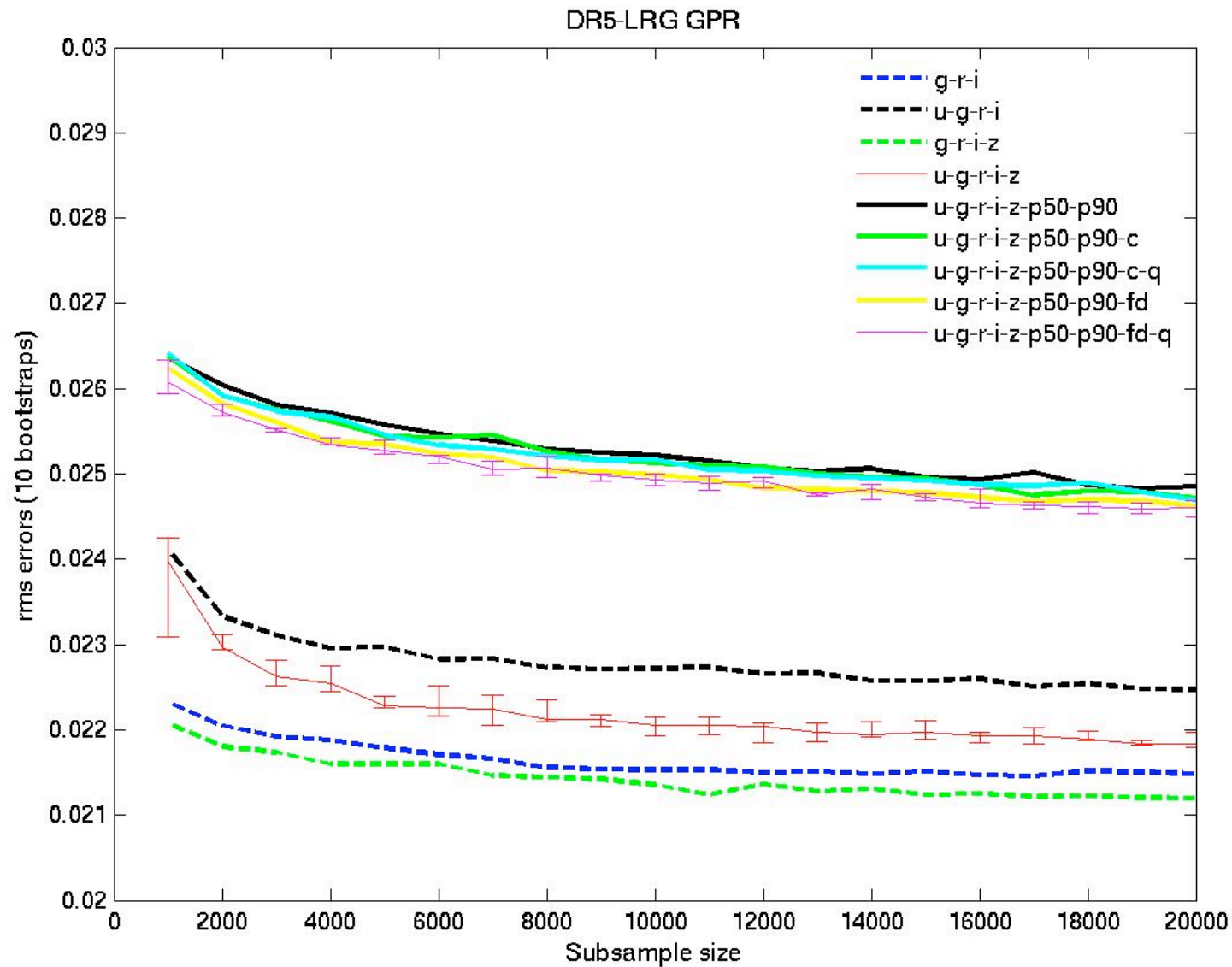
SDSS-DR5/LRG GPR



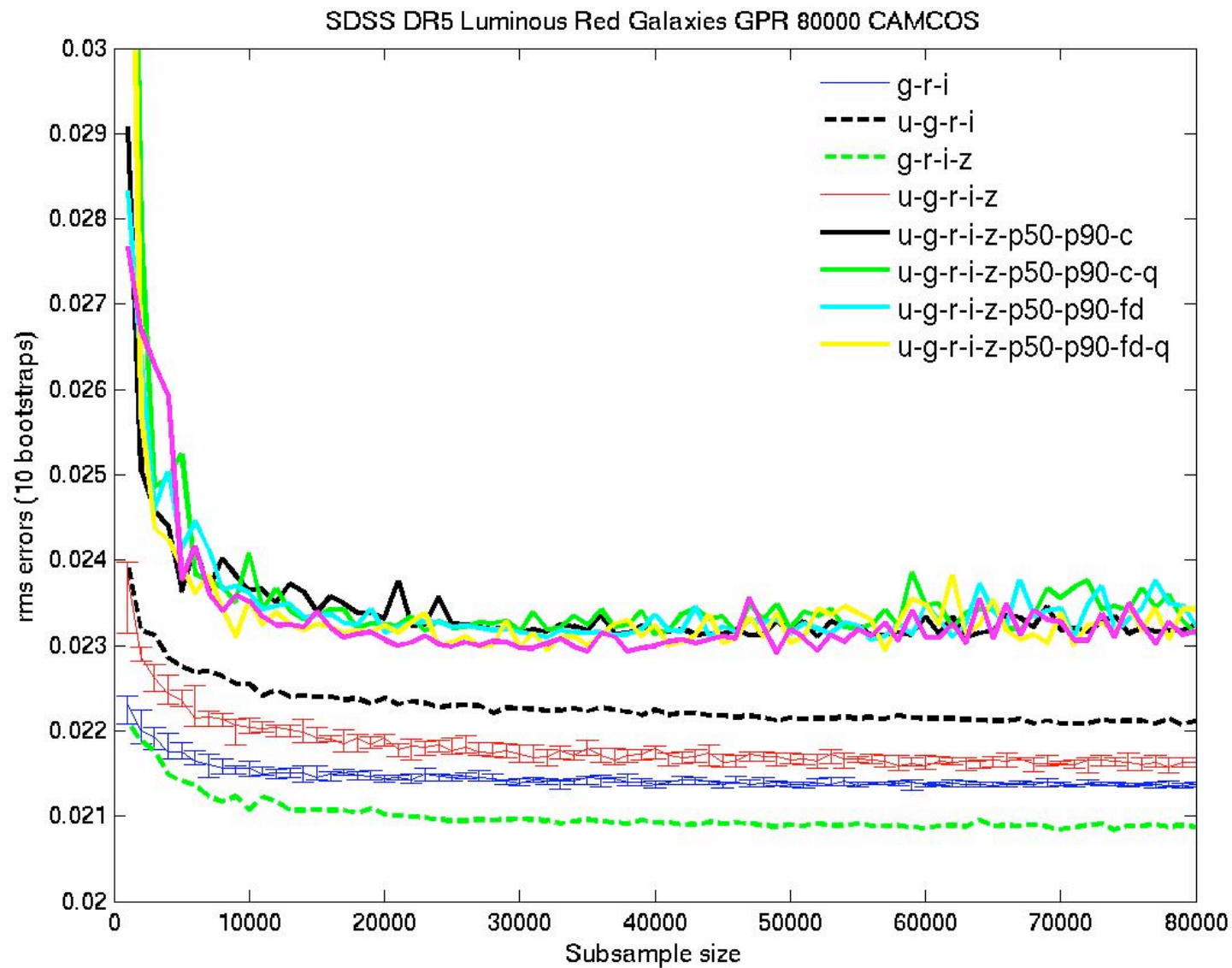
SDSS-DR5/LRG GPR



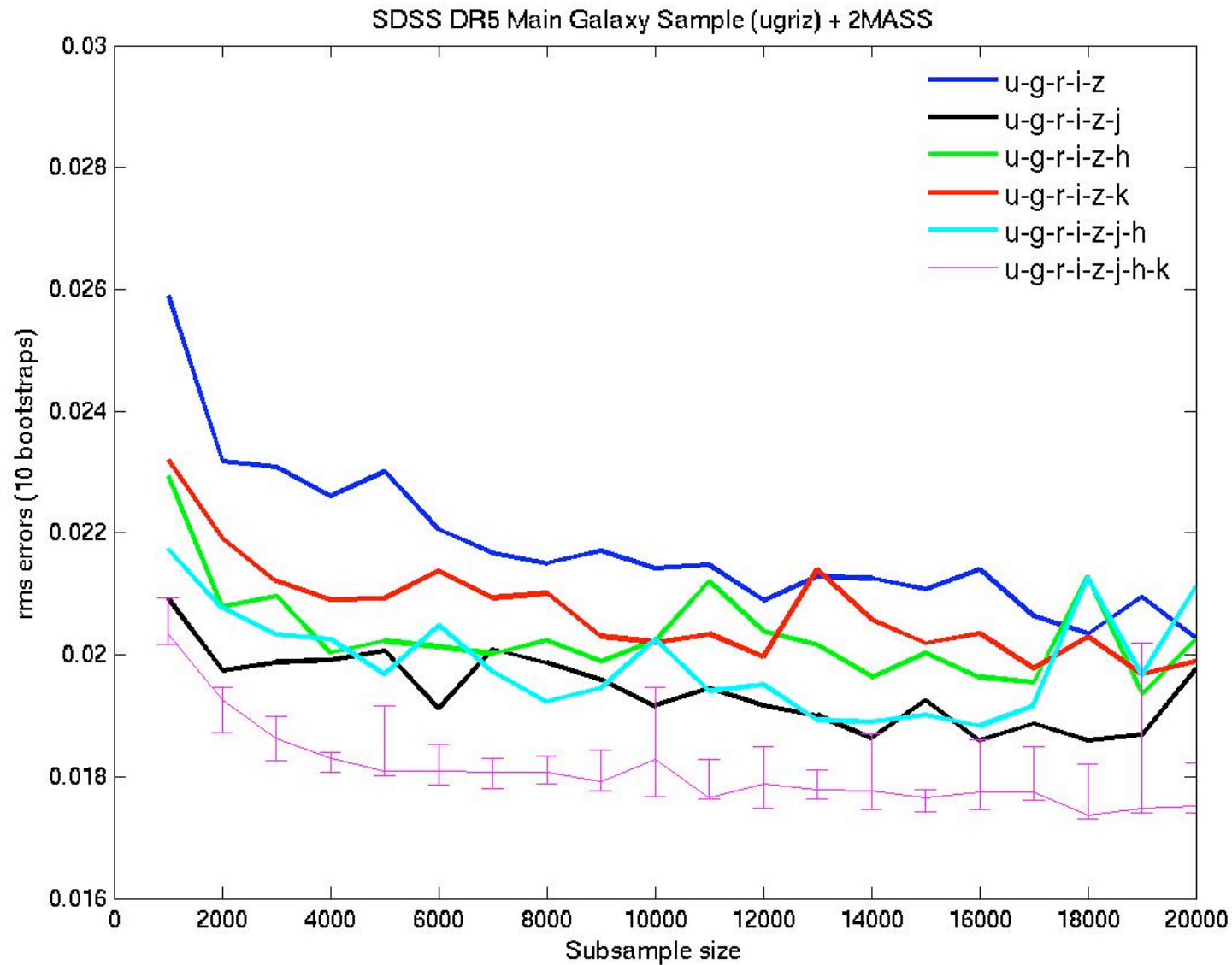
Luminous Red Galaxies 20000 GPR



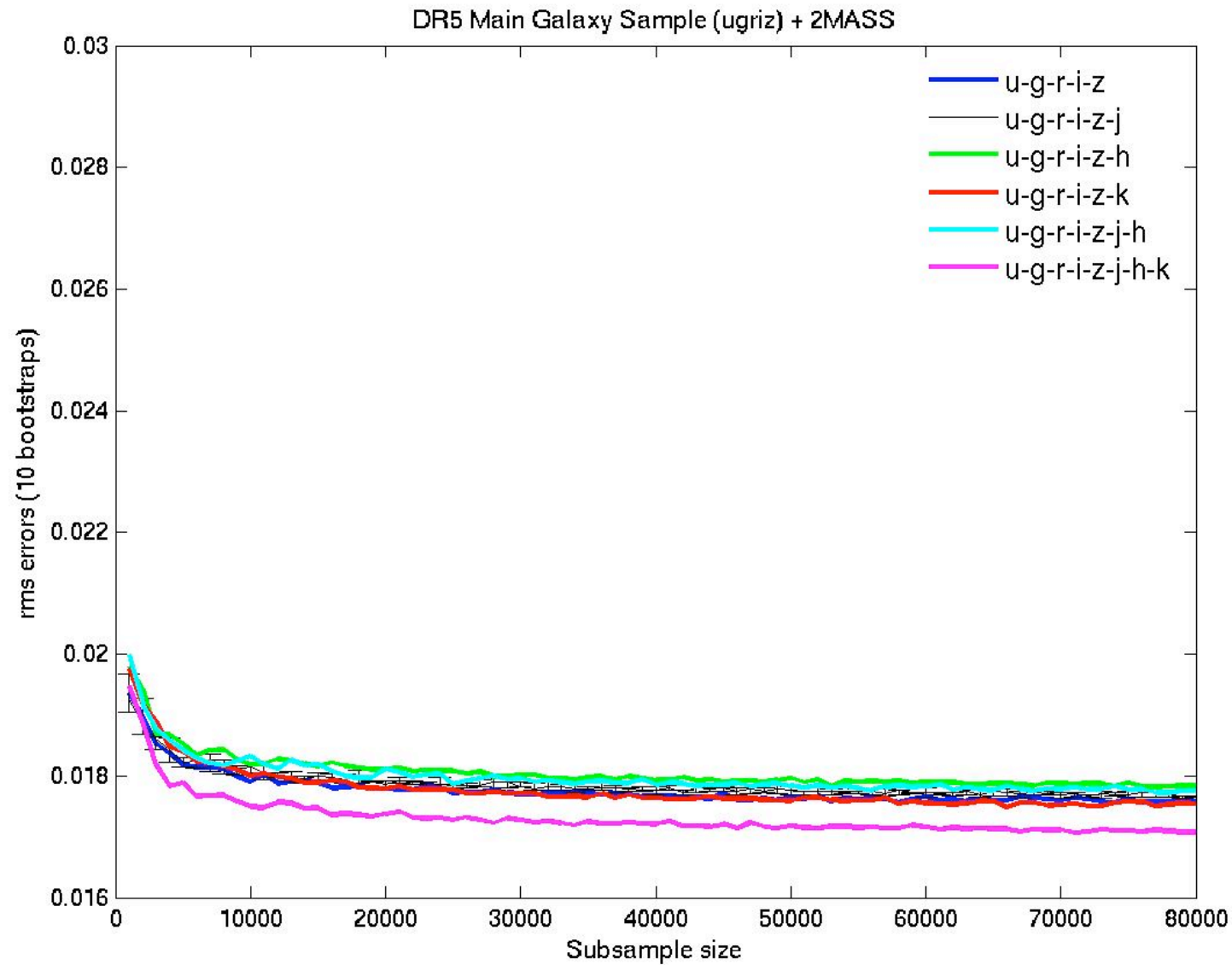
LRG 80000 GPR



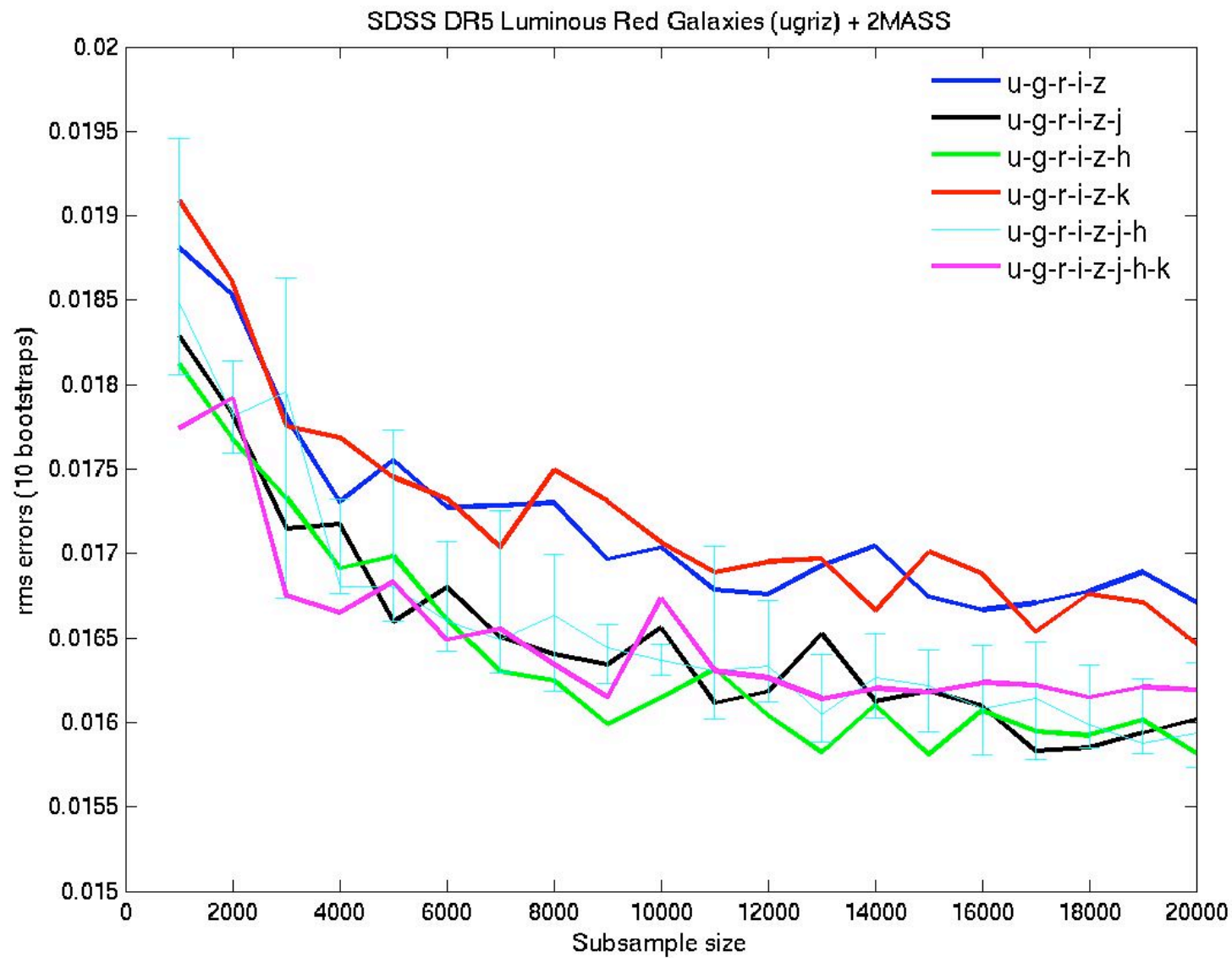
MGS + 2MASS 20000



MGS + 2MASS 80000



LRG+ 2MASS 20000



LRG+ 2MASS 80000

